Análisis Multiresolución aplicado a Segmentación Fonética Independiente del Texto

Analía S. Cherniz[†] [‡], María E. Torres[†] [§], Hugo L. Rufiner[‡] [§] y Anna Esposito[♭]
[†]*Laboratorio de Señales y Dinámicas no Lineales y* [‡]*Laboratorio de Cibernética.*
Facultad de Ingeniería, Universidad Nacional de Entre Ríos,
C.C. 47 Suc. 3 - 3100 Paraná (E.R.), Argentina
[§]*Laboratorio de Señales e Inteligencia Computacional*
Universidad Nacional del Litoral, Santa Fe, Argentina
[♭]*Department of Psychology and IIASS*
Second University of Naples, Caserta, Italy
metorres@ceride.gov.ar

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
A. S. Cherniz, M. E. Torres, H. L. Rufiner & A. Esposito; "Analisis Multirresolución aplicado a la segmentacion fonética independiente del texto"
Revista Argentina de Bioingeniería, Vol. 13, No. 2, pp. 18-23, oct, 2007.

***Resumen*— La segmentación automática del habla es importante en distintas aplicaciones. Los métodos utilizados comunmente se basan en modelos ocultos de Markov. Estos modelan estadísticamente las unidades fonéticas y realizan una alineación forzada de los datos según una transcripción conocida. Este proceso es costoso y consume tiempo debido a la gran cantidad de datos necesarios para entrenar el sistema. Como solución se han propuesto procedimientos de segmentación independientes del texto. Estos detectan transiciones en la evolución de los parámetros que representan la señal de habla. En estos procedimientos la forma de representar o parametrizar la señal juega un rol central. En este trabajo se proponen nuevas parametrizaciones basadas en la entropía multiresolución continua, utilizando entropía Shannon, y en la divergencia multiresolución continua, mediante la distancia Kullback-Leibler. Dichas propuestas se comparan con la parametrización Melbank clásica. Los resultados muestran que el desempeño del algoritmo de segmentación se incrementa con estas alternativas. En particular, la parametrización basada en la divergencia multiresolución continua muestra los mejores resultados, incrementando el número de límites correctamente detectados y disminuyendo la cantidad de puntos insertados erroneamente. Esto sugiere que estas parametrizaciones proveen una mejor información, relacionada con características acústicas del habla vinculadas a las transiciones fonémicas.**

***Palabras clave*— Entropía Shannon; Distancia Kullback-Leibler; Análisis multiresolución; Segmentación automática del habla.**

I. Introducción

La segmentación y etiquetado de las señales de habla, según reglas fonéticas o lingüísticas, es fundamental en distintas aplicaciones en el campo del procesamiento del habla. El objetivo de esta tarea es organizar la secuencia obtenida a partir del análisis por tramos de la señal de habla en segmentos homogéneos asociados con fonemas, palabras, sílabas u otras unidades acústicas. Tradicionalmente, esto ha sido realizado manualmente por expertos fonetistas, utilizando información visual y auditiva. Sin embargo, este procedimiento puede ser tedioso, subjetivo, consumir mucho tiempo y ser propenso a errores, especialmente para registros de habla espontánea [1].

Los métodos de segmentación automática más comunes modelan estadísticamente las transiciones entre las unidades fonéticas mediante modelos ocultos de Markov (HMMs) [2]. Los HMMs se entrenan utilizando una base de datos con sus correspondientes transcripciones y son usados luego para hacer un alineamiento forzado de los datos conforme a una trascripción fonética conocida. Este proceso es costoso debido a la gran cantidad de datos necesarios para entrenar los modelos acústicos.

Distintos métodos de segmentación independientes del texto han sido sugeridos para solucionar alguno de estos inconvenientes [3]–[5]. En [3] los autores proponen un algoritmo que trabaja sobre un número arbitrario de parámetros, obtenidos a través del análisis por tramos de la señal de habla. Este procedimiento intenta detectar transiciones en los segmentos de la señal donde los valores de los parámetros cambian de forma rápida y significativa.

Nociones de entropía han sido utilizadas para caracterizar el grado de complejidad del habla y otras señales fisiológicas [6]. La entropía espectral ha sido utilizada de diversas maneras para tareas de segmentación de sentencias o palabras y detección de silencios [7], [8]. La entropía multiresolución continua (CME) da cuenta de la evolución temporal de las entropías Shannon o Tsallis calculadas sobre los coeficientes de la transformada ondita contínua (CWT) [9]. Recientemente, la CME ha sido incluída en la parametrización del habla de un sistema de reconocimiento automático, mejorando su desempeño [10].

En este trabajo se propone utilizar una parametrización del habla basada en la CME, usando entropía Shannon, y una basada en la divergencia multiresolución contínua (CMD), usando la distancia de Kullback-Leibler, como entradas del método de segmentación automática del habla propuesto en [3]. Los resultados obtenidos con las codificaciones propuestas son comparados con la segmentación lograda usando la parametrización clásica o Melbank.

Este trabajo es financiado por la A.N.P.C.yT., bajo el Proyecto PICT Nro 11-12700, Argentina.

II. Materiales y métodos

A. Parametrización Melbank

La parametrización en bancos de filtros en frecuencia de mel (Melbank) es un procesamiento por tramos estándar de la señal de habla, que mide la energía de señal en una determinada banda de frecuencia [11]. Los filtros se distribuyen según una escala que intenta reproducir la resolución espectral del oído humano:

$$Mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right), \qquad (1)$$

En [3] se utilizaron diferentes parametrizaciones estándar, con distinta cantidad de parámetros, siendo Melbank con 8 coeficientes la codificación que mostró mejores resultados.

B. Parametrización basada en CME

Dada la señal discreta de habla $\mathbf{s} = \{s[k], k = 1, \dots K\}$, de longitud K, para obtener la CME se calcula primero la transformada ondita cuasi–continua $\Psi_{\mathbf{s}}(a, b)$. Esto conduce a una descomposición discretizada $\{d[j,k]\} = \{\Psi_{\mathbf{s}}(a = j\delta, b = k)\}$ en el plano tiempo–escala, donde $j = 1, \dots, J \in \mathbb{Z}$, $\delta \in \mathbb{R}^+$ y $b = k$ es el control temporal. Para j fijo, la evolución de los coeficientes de la CWT se denotarán como $\mathbf{d}_j = \{d_j[k]\}$.

Se considera el conjunto de ventanas deslizantes $W^j(m, L, \Delta) = \{d_j[k], k = l + m\Delta,\ l = 1, ..., L\}$, para $m = 0, 1, ..., M$, donde $L \in \mathbb{N}$ es el ancho de las mismas y $\Delta \in \mathbb{N}$ su desplazamiento. Estos parámetros se eligen tal que $L \leq K$ y $(K - L)/\Delta = M \in \mathbb{Z}$. Esta selección esta directamente relacionada con la máxima velocidad de modificación del tracto vocal [12].

Cada ventana se divide en N subintervalos disjuntos I_n y se denota con $p_{j,m}(I_n)$ la probabilidad de que un dado $d_j[k] \in W_j(m, L, \Delta)$ pertenezca a uno de estos subintervalos. Para cada ventana se obtiene el siguiente conjunto de probabilidades:

$$\{P[j,m]\} = \{p_{j,m}(I_n),\ n = 1, \dots, N\}. \qquad (2)$$

La entropía Shannon de (2) es la siguiente:

$$\mathcal{H}_{\mathbf{d}}[j,m] = -\sum_{n=1}^{N} p_m^j(I_n^j) \ln\left(p_m^j(I_n^j)\right). \qquad (3)$$

La matriz $\mathbf{CME} = \{\mathcal{H}_{\mathbf{d}}[j,m]\}$ es la entropía multiresolución continua de $\mathbf{s}$, para cada escala j y segmento m.

Se utiliza el análisis de componentes principales (PCA) para extraer las características de mayor varianza y obtener la nueva parametrización. A partir de $\mathbf{U} = \mathbf{CME}^*$ (matriz $\mathbf{CME}$ normalizada estadísticamente) se calcula la matriz de correlación $\sigma = \mathbf{U}\mathbf{U}^T$. Se obtienen las matrices de los eigenvectores $\mathbf{Q}$, con sus eigenvalores $\mathbf{\Lambda}$ asociados, tal que $\sigma = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$. La matriz de componentes principales es:

$$\mathbf{Y} = \mathbf{Q}^T\mathbf{U}. \qquad (4)$$

La fila $\mathbf{y}_1 = \{y_1[m], m = 0, 1, ..., M\}$ es el componente principal de $\mathbf{Y}$ y corresponde al valor máximo de $\mathbf{\Lambda}$. Se propone una nueva parametrización usando las filas de $\mathbf{Y}$ asociadas a los $\mathcal{J}$ mayores valores de $\mathbf{\Lambda}$: $\mathbf{y}_i$, $i = 1, .., \mathcal{J}$, con $\mathcal{J} = 8$. Se elige esta cantidad de componentes en concordancia con la parametrización Melbank utilizada en la comparación de resultados. Además, las mismas acumulan más del 95 % de la variabilidad total de la señal.

C. Parametrización basada en CMD

Teniendo en cuenta el conjunto de probabilidades (2), para la ventana $W^j(m, L, \Delta)$, se considera ahora $\{R^j[m]\} = \{r_m^j(I_n^j),\ n = 1, \dots, N\}$, un segundo conjunto correspondiente a $W^j(m + 1, L, \Delta)$, la ventana siguiente. La divergencia de Kullback–Leiber para estas ventanas es:

$$\mathcal{D}_{\mathbf{d}}[j,m] = \sum_{n=1}^{N} p_m^j(I_n^j) \ln\left(\frac{p_m^j(I_n^j)}{r_m^j(I_n^j)}\right) \qquad (5)$$

Calculando esto para todas las ventanas y escalas se obtiene la matriz de la divergencia multiresolución continua $\mathbf{CMD} = \{\mathcal{D}_{\mathbf{d}}[j,m]\}$.

Se aplica PCA para obtener la parametrización basada en CMD. Esto se realiza de la misma manera que para el caso de la codificación basada en CME. Se obtiene nuevamente, un conjunto de $\mathcal{J} = 8$ parámetros para cada segmento m: $\mathbf{y}_i = \{y_i[m], m = 0, ..., M\}$, $i = 1, .., \mathcal{J}$.

D. Algoritmo de segmentación del habla

Se utilizan las codificaciones anteriores para realizar la segmentación mediante el algoritmo propuesto en [3]. Este es regulado por tres parámetros operacionales: α, β y γ.

El parámetro α identifica cuántos segmentos consecutivos se necesitan para estimar la intensidad de un cambio abrupto. Así, dado $\{y_i[m]\}$, se calcula la siguiente función:

$$\mathcal{F}_i^\alpha[m] = \left| \sum_{\mu=m-\alpha}^{m-1} \frac{y_i[\mu]}{\alpha} - \sum_{\mu=m+1}^{m+\alpha} \frac{y_i[\mu]}{\alpha} \right|. \qquad (6)$$

Se utiliza un procedimiento de umbralamiento relativo para identificar el segmento m^* donde un pico, relacionado a una posible transición fonética, es detectado de acuerdo al parámetro β. Dado un intervalo $[u, v] \in [\alpha, M - \alpha]$, donde $\mathcal{F}_i^\alpha[u]$ y $\mathcal{F}_i^\alpha[v]$ son dos valles de (6), el segmento $m^* \in [u, v]$ se selecciona de forma tal que $\mathcal{F}_i^\alpha[m^*]$ es un máximo local en dicho intervalo. Se calcula la altura relativa:

$$\eta = \min\left[\mathcal{F}_i^\alpha[m^*] - \mathcal{F}_i^\alpha[u], \mathcal{F}_i^\alpha[m^*] - \mathcal{F}_i^\alpha[v]\right]. \qquad (7)$$

Se considera la matriz $\mathbf{T}$, donde $T[i,m] = 1$ si $\eta \geq \beta$, para la secuencia i en el intervalo m, y 0 en otro caso.

Las transiciones detectadas por distintas características i no ocurren simultáneamente, sino en un corto intervalo de tiempo. La segmentación usa un procedimiento de ajuste para combinar en un baricentro los eventos de cada grupo de transiciones abruptas cuasi-simultáneas. El parámetro γ identifica el ancho del vecindario donde se individualiza el baricentro. Para $m = 1, ..., M - \gamma + 1$ se considera el intervalo $V = [m, m + \gamma - 1] \in \mathbb{N}$ y se calcula la función:

$$G[c] = \sum_{\mu=c}^{c+\gamma-1} \sum_{i=1}^{\mathcal{J}} T(i,\mu)\,|\mu - c|\,, \quad c \in V. \qquad (8)$$

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
A. S. Cherniz, M. E. Torres, H. L. Rufiner & A. Esposito; "Analisis Multirresolución aplicado a la segmentacion fonética independiente del texto"
Revista Argentina de Bioingeniería, Vol. 13, No. 2, pp. 18-23, oct, 2007.

El posible baricentro en V es el segmento $\widetilde{c}$ donde $G[\widetilde{c}] = \min_{V} G[c]$. El valor $\widetilde{G}[m]$ indica cuántos baricentros $\widetilde{c}$ se han encontrado en m. Esto produce una función cuyos picos corresponden a la indicación de un posible límite fonético.

E. Señales y base de datos

Se utilizó un subconjunto del corpus de habla Albayzin de 600 sentencias relacionadas con geografía española, con un vocabulario de 200 palabras. Las frases tienen una duración promedio de 3.55 s. y fueron pronunciadas por 6 hombres y 6 mujeres (edad promedio: 31.8 años) de la región central de España. Como referencia se utilizaron archivos de habla etiquetada, usando segmentación automática asistida por expertos, que registran la posición de los límites fonéticos expresadas en ms. Cada frase ha sido normalizada en media, pre-enfatizada y particionada mediante ventanas de Hamming en segmentos de 20 ms, desplazados 10 ms [12]. Las tres parametrizaciones se codificaron con 8 coeficientes.

F. Índices para evaluar el desempeño de segmentación

Para evaluar la segmentación se calcularon el porcentaje de límites fonéticos detectados correctamente (PC) y el porcentaje de puntos erroneamente insertados (PI).

El índice PC relaciona la cantidad de límites correctamente detectados, B_C, con el total de límites fonéticos de la base de datos, B_T, usando una tolerancia de ± 20 ms.

$$PC = 100\frac{B_C}{B_T}. \tag{9}$$

El índice PI relaciona el número de límites fonéticos detectados erroneamente, $B_I = B_D - B_C$ (B_D: cantidad total de puntos de segmentación detectados), con el número total de segmentos F_T presentes en la señal.

$$PI = 100\left(\frac{B_I}{F_T}\right), \tag{10}$$

Para evaluar la significancia estadística de los resultados se estimó la probabilidad de que las codificaciones propuestas sean mejores que la parametrización Melbank ($\Pr(\epsilon < \epsilon_{ref})$). Para ello se supuso la independencia estadística de los errores de detección y se aproximó la distribución binomial de los mismos por medio de una Gaussiana.

III. RESULTADOS

La Tabla I muestra los índices PC y PI para los esquemas de codificación propuestos (Secs. II.B y II.C) los cuales se comparan con la parametrización Melbank, usando los siguientes parámetros operacionales para el algoritmo de segmentación: $\gamma = 3$, $\alpha = 2$, 4, 6 y $\beta = 0{,}01$, 0,05, 0,1. Los números en negrita indican el mejor PC obtenido.

Comparando los índices en negrita, la codificación basada en CME mostró el mejor PC, pero su índice PI fue alto. Por el contrario, la codificación basada en la CMD mostró un PC aceptable, con menor PI que la parametrización Melbank. Es importante notar que los índices PC y PI óptimos son aquellos que dan 100 % y 0 % respectivamente.

TABLA I

PC Y PI DEL ALGORITMO DE SEGMENTACIÓN PARA LAS TRES CODIFICACIONES EVALUADAS UTILIZANDO DIFERENTES PARÁMETROS OPERACIONALES. EN NEGRITA SE INDICAN LOS MEJORES RESULTADOS.

Parámetros			Codificaciones					
			Melbank		CME		CMD	
γ	α	β	PC	PI	PC	PI	PC	PI
3	2	0.01	**86.52**	16.63	81.64	16.79	84.65	15.12
		0.05	83.97	15.37	84.40	17.44	81.80	8.36
		0.1	79.97	10.49	85.48	17.51	77.20	6.96
	4	0.01	81.11	13.56	83.79	14.90	83.09	12.59
		0.05	78.78	11.96	83.65	15.00	79.31	7.01
		0.1	75.92	7.24	83.39	14.83	74.98	6.08
	6	0.01	75.41	13.31	86.25	17.43	**86.75**	13.87
		0.05	72.02	8.77	**86.91**	17.16	83.51	9.04
		0.1	68.05	5.22	86.09	16.61	79.94	8.05

La evaluación de la significancia estadística de estos resultados para el índice PC mostró que $\Pr(\epsilon < \epsilon_{ref}) > 93{,}02\,\%$ para la parametrización basada en CME y $\Pr(\epsilon < \epsilon_{ref}) > 80{,}57\,\%$ para la basada en CMD. El índice PI para la codificación por CME se comportó peor que para Melbank con $\Pr(\epsilon > \epsilon_{ref}) > 96{,}11\,\%$. El PI correspondiente a la parametrización basada en CMD fue significativamente menor, con $\Pr(\epsilon < \epsilon_{ref}) > 99{,}99\,\%$ respecto de Melbank.

Se observa, en general, que el índice PC para las parametrizaciones basadas en CME y CMD es mayor que el de Melbank. Con respecto al índice PI, se puede ver que la parametrización basada en CMD es la que presenta los valores más bajos, lo cual supone un mejor desempeño del algoritmo con esta codificación.

En la Fig. 1 se compara el desempeño del algoritmo de segmentación para las codificaciones propuestas y Melbank, utilizando $\alpha = 2, 3, 4, 5, 6$, con $\gamma = 3$ y $\beta = 0{,}01$. La Fig. 1(a) muestra el índice PC para los distintos valores de α. Cuanto más se acerca la línea al límite superior mejor es el desempeño de la segmentación. Se observa que las codificaciones propuestas muestran un mejor comportamiento cuando α se incrementa, mientras que Melbank decae. En la Fig. 1(b) se observa el índice PI para estos experimentos. La línea más cercana al límite inferior es la que presenta mejores resultados. En este caso, la codificación basada en la CMD muestra el mejor desempeño. Si bien cuando $\alpha = 5$ y 6 Melbank se comporta mejor, los PI de la misma son comparables a los de la CMD.

Los resultados de la segmentación usando las codificaciones propuestas parecen más estables que Melbank ante cambios del parámetro α. Esto puede deberse a que, mientras los cambios en la evolución de los parámetros Melbank son suaves y aparecen reflejados en varios segmentos de la señal de habla, la evolución de los coeficientes para la CME o la CMD es más abrupta y concentrada. Por lo tanto, la cantidad de segmentos utilizados para la detección, determinado por α, afecta mayormente a la parametrización Melbank.

Esta característica de la CME, que permite detectar cambios en los parámetros de manera concentrada, provoca también que la codificación basada en la misma tenga altos

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
A. S. Cherniz, M. E. Torres, H. L. Rufiner & A. Esposito; "Analisis Multirresolución aplicado a la segmentacion fonética independiente del texto"
Revista Argentina de Bioingeniería, Vol. 13, No. 2, pp. 18-23, oct, 2007.

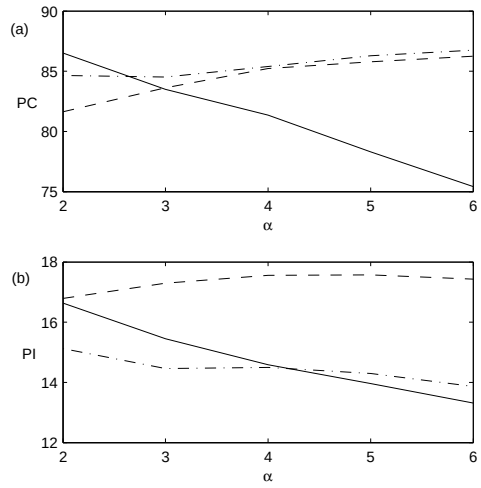

Fig. 1. Porcentajes de (a) límites fonéticos detectados correctamente PC y (b) puntos erróneamente insertados PI, obtenidos para el algoritmo de segmentación usando la parametrización Melbank clásica (línea llena) y la codificaciones basadas en CME (línea entrecortada) y CMD (línea entrecortada con puntos), para $\alpha = 2, 3, 4, 5, 6$, $\beta = 0{,}01$ y $\gamma = 3$.

valores de PI. Esto se debe a que esta herramienta realza los pequeños cambios que se producen segmento a segmento, lo cual provoca que aparezcan muchos falsos positivos. En la CMD, en cambio, como se tiene en cuenta la diferencia que existe entre segmentos, la codificación es más robusta.

La Fig. 2 muestra en (a) el índice PC y en (b) el PI para las tres codificaciones evaluadas, usando $\gamma = 3, 4, 5, 6, 7$, $\alpha = 2$ y $\beta = 0{,}01$. El mejor desempeño se observa para la parametrización basada en CMD, ya que presenta buenos valores de PC, especialmente para $\gamma = 4, 6$ y 7, y la curva para el índice PI es la menor para todos los valores de γ.

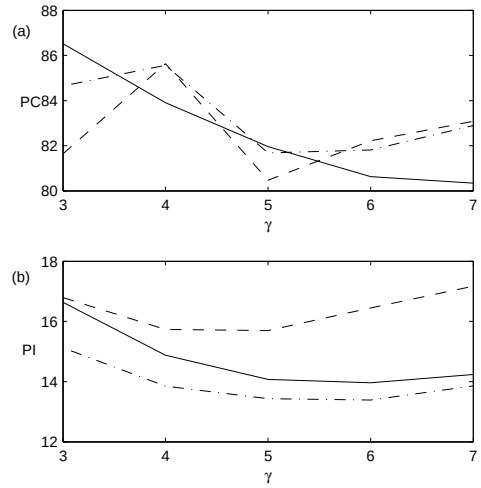

Fig. 2. Porcentajes de (a) límites fonéticos detectados correctamente PC y (b) puntos erróneamente insertados PI, obtenidos para el algoritmo de segmentación utilizando la parametrización Melbank (línea llena) y la codificaciones basadas en CME (línea entrecortada) y CMD (línea entrecortada con puntos), para $\gamma = 3, 4, 5, 6, 7$, $\alpha = 2$ y $\beta = 0{,}01$.

Se puede ver que las curvas de la Fig. 2(b) tienen forma cóncava, con mínimos alrededor de $\gamma = 5$. Este parámetro determina el vecindario de segmentos utilizados para realizar el procedimiento de ajuste. Valores bajos de γ producen vecindarios angostos que pueden no cubrir todo el rango donde se manifiesta la transición fonética. Esto puede provocar que un cambio se considere como dos transiciones, aumentando el índice PI. Por otro lado, cuando γ es grande, dos transiciones que se manifiesten en segmentos cercanos pueden ser interpretadas como un único punto, haciendo que el desempeño del segmentador se deteriore. Esto también se observa en indice PC, en especial para Melbank.

IV. Conclusiones

Los resultados indican que las dos parametrizaciones propuestas mejoran el desempeño del algoritmo de segmentación utilizado. En particular, la parametrización basada en CMD muestra los mejores resultados, ya que incrementa el número de límites fonémicos detectados correctamente disminuyendo la cantidad de puntos erroneamente insertados. Esto sugiere que estas nuevas codificaciones proveen información valiosa para el algoritmo de segmentación, relacionada con características acústicas vinculadas a transiciones fonéticas. Las medidas de información utilizadas podrían reflejar entonces cambios en la dinámica del tracto vocal, lo cual es un dato importante para realizar la segmentación.

Referencias

[1] D. Binnenpoorte, S. Goddijn, y C. Cucchiarini, "How to improve human and machine transcriptions of spontaneous speech," en *ISCA/IEEE Workshop on Spont. Speech Proc. Recog.*, pp. 147–150, 2003.

[2] D. Torre, L. Hernández, y L. Villarrubia, "Automatic phonetic segmentation," *IEEE Transations on Speech and Audio Processing*, vol. 11, no. 6, pp. 617–625, 2003.

[3] A. Esposito y G. Aversano, "Text independent methods for speech segmentation," en *Nonlinear Speech Modeling And Applications: Advanced Lectures and Revised Selected Papers*. Springer, 2005, pp. 261–290.

[4] J. Gómez y M. Castro, "Automatic segmentation of speech at the phonetic level," en *Structural, Syntactic, and Statistical Pattern Recognition*. London, UK: Springer Berlin / Heidelberg, p. 672.

[5] M. Sharma y R. Mammone, "Blind speech segmentation: Automatic segmentation of speech without linguistic knowledge," en *Proc. of Fourth Int. Conf. on Spoken Language*, pp. 1237–1240. IEEE, 1996.

[6] H. Rufiner, M. Torres, L. Gamero, y D. Milone, "Introducing complexity measures in nonlinear physiological signal: application to robust speech recognition," *Physica A*, vol. 332, pp. 496–508, 2004.

[7] B.-F. Wu y K.-C. Wang, "Noise spectrum estimation with entropy-based VAD in non-stationary environments," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E89-A, no. 2, pp. 479–485, 2006.

[8] K. Weaver, K. Waheed, y F. Salem, "An entropy based robust speech boundary detection algorithm for realistic noisy environments," en *Proc. of the International Joint Conference on Neural Networks*, pp. 680–685, 2003.

[9] M. Torres, L. Gamero, P. Flandrin, y P. Abry, "On a multiresolution entropy measure," en *SPIE'97 Wavelet Applications in Signal and Image Processing V*, pp. 400–407, 1997.

[10] M. Torres, H. Rufiner, D. Milone, y A. Cherniz, "Comparison between temporal and time-scale information measures applied to speech recognition," *WSEAS Transactions on signal Processing*, vol. 9, no. 2, pp. 1153–1159, 2006.

[11] L. Rabiner y B. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, New Jersey: Prentice-Hall, 1993.

[12] J. Deller, J. Proakis, y J. Hansen, *Discrete Time Processing of Speech Signals*. New York: Macmillan Publishing, 1993.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
A. S. Cherniz, M. E. Torres, H. L. Rufiner & A. Esposito; "Analisis Multirresolución aplicado a la segmentacion fonética independiente del texto"
Revista Argentina de Bioingeniería, Vol. 13, No. 2, pp. 18-23, oct, 2007.